The Role of Big Data Analytics in Predicting Power System Outages

Mladen Kezunovic
Regents Professor
Texas A&M University

Pserc Webinar
February 18, 2020

Outline

The Importance of Outages: Huge impact

The Value of Prediction: Mitigation

The Power of Data Analytics: Risk assessment

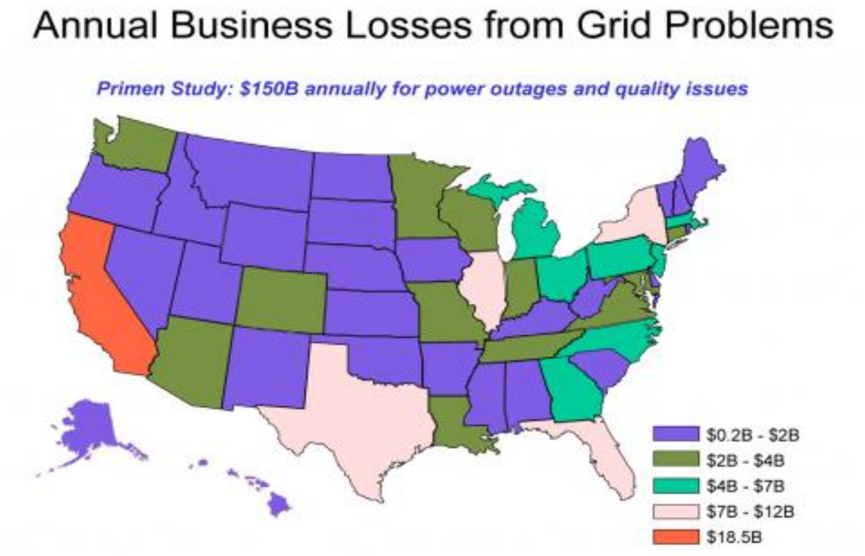
Overall Goal: Improved resilience

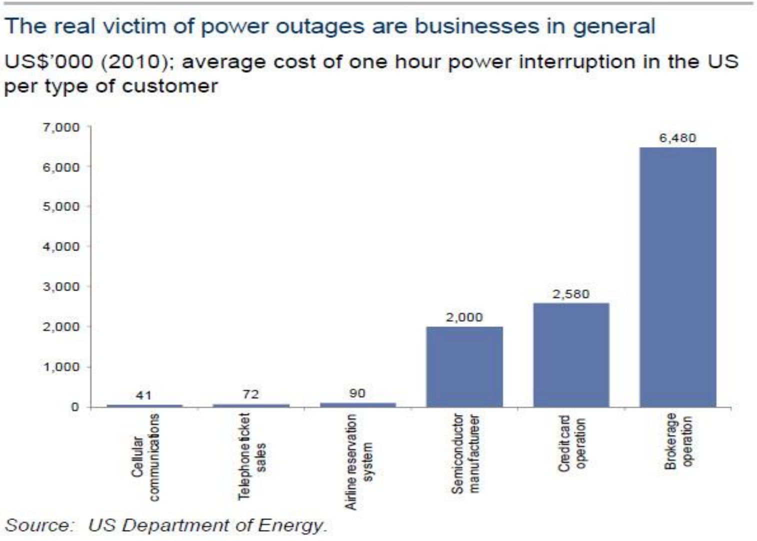
Big Data: Properties

Big Data Analytics: Challenges

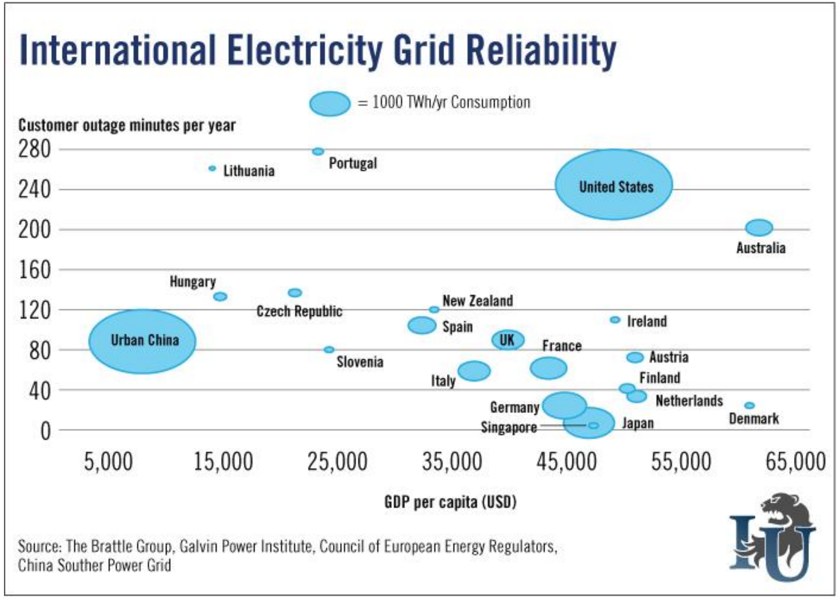
Applications: Outages caused by weather

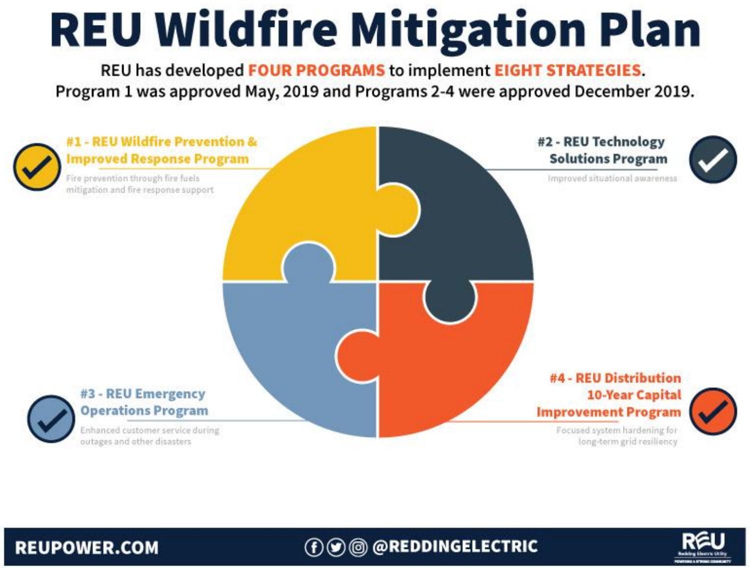
The Importance of Outages: Huge impact





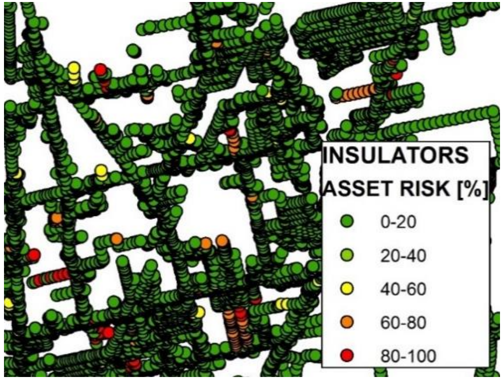
The Value of Prediction: Mitigation



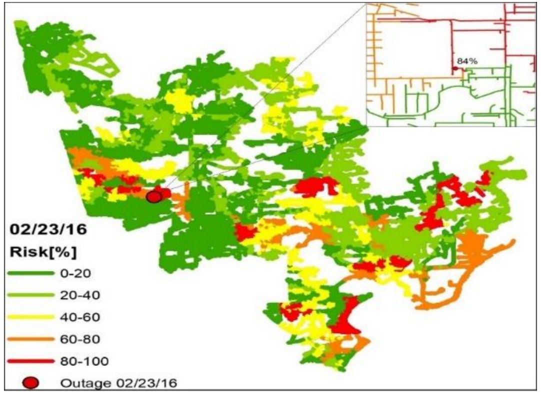


The Power of Data Analytics: Risk assessment

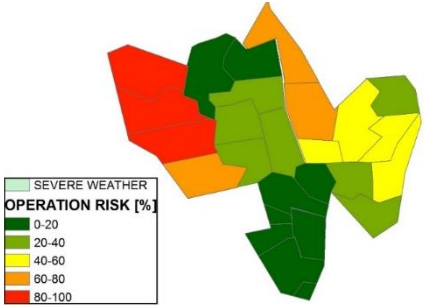
Asset Management Risk Map



Outage Management Risk Map

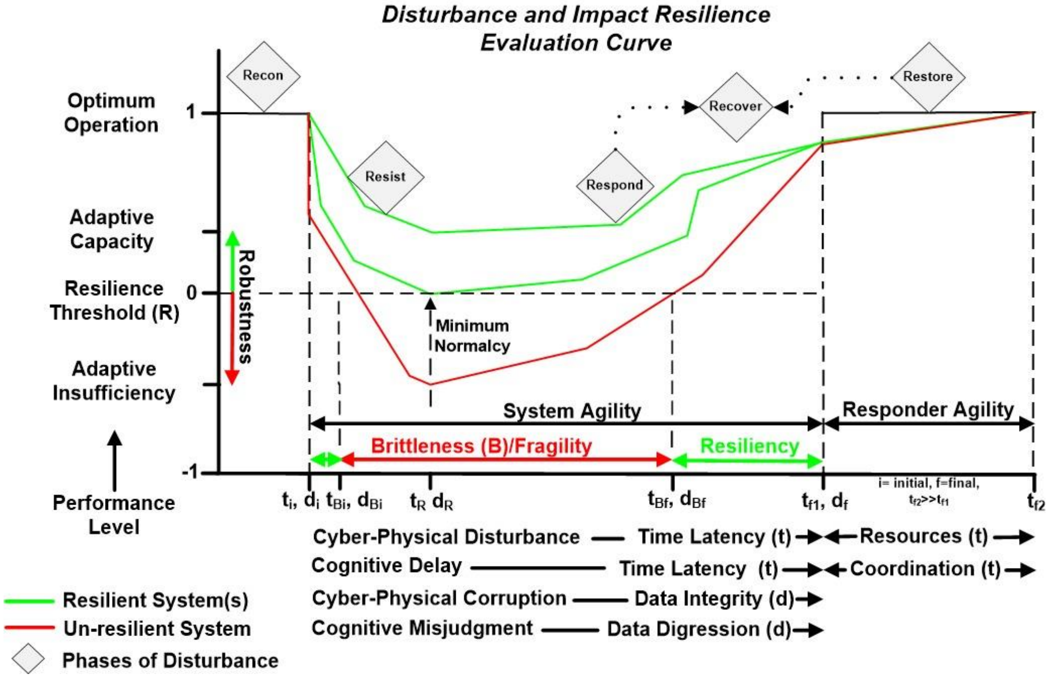


Operation Risk Map



M. Kezunovic, Z. Obradovic, T. Dokic, B. Zhang, J. Stojanovic, P. Dehghanian, and P. -C. Chen, "Predicating Spatiotemporal Impacts of Weather on Power Systems using Big Data Science," Springer Verlag, Data Science and Big Data: An Environment of Computational Intelligence, Pedrycz, Witold, Chen, Shyi-Ming (Eds.), ISBN 978-3-319-53474-9, 2017.

Overall Goal: Improved resilience

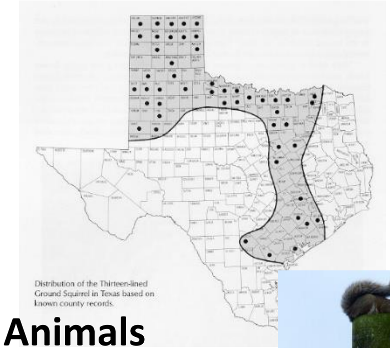


K. Eshghi, B.K. Johnson, C. G Rieger "Power System Protection and Resiliency Metrics." 2015 Resilience Week, Workshop Proceedings, Idaho National Laboratory, August 2015

T. Mc. Junkin, C.G. Rieger, "Electricity distribution system resilient control system metrics." 2017 Resilience Week, Workshop Proceedings, Idaho National Laboratory, September 2017

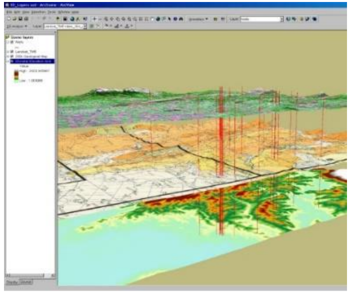
Big Data: Properties

Utility measurements

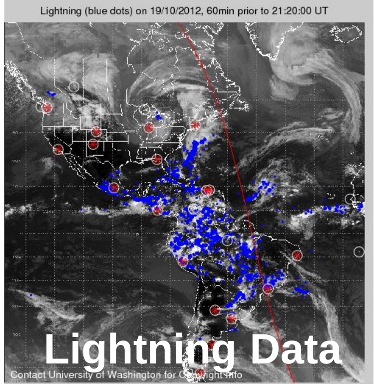


Animals Data

Market data

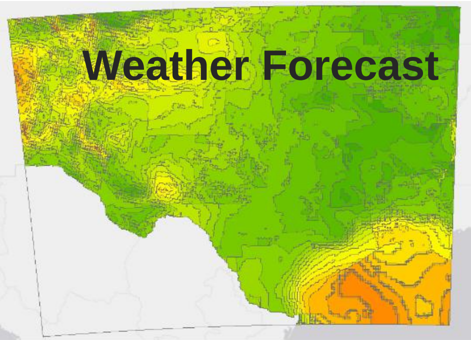


GIS

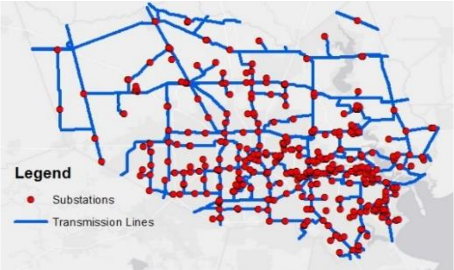


Lightning Data

UAS



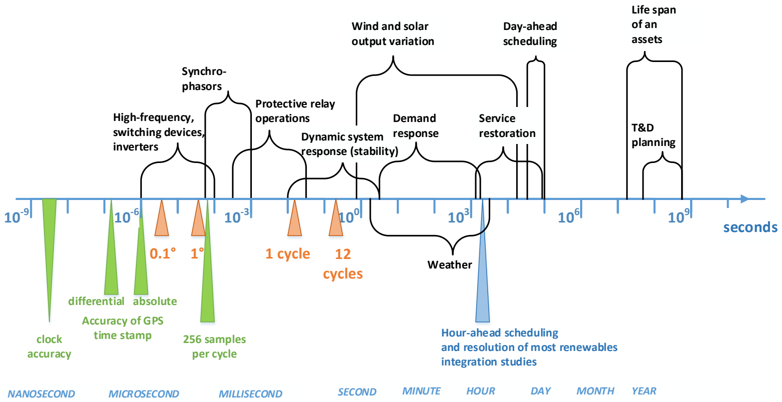
Weather Forecast

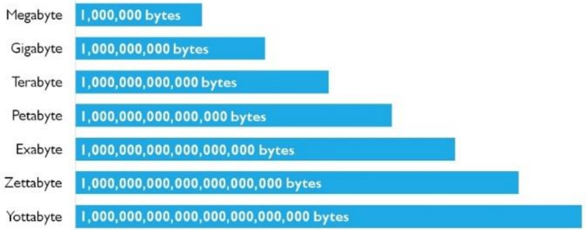


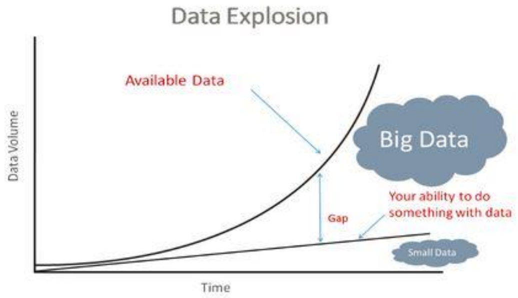
Network Assets Data

Vegetation Indices

Big Data: Properties

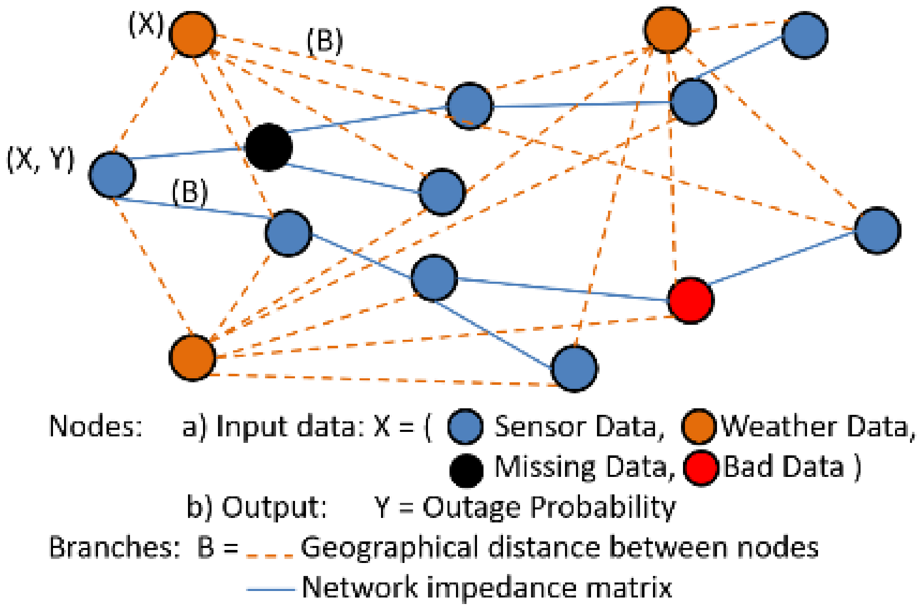




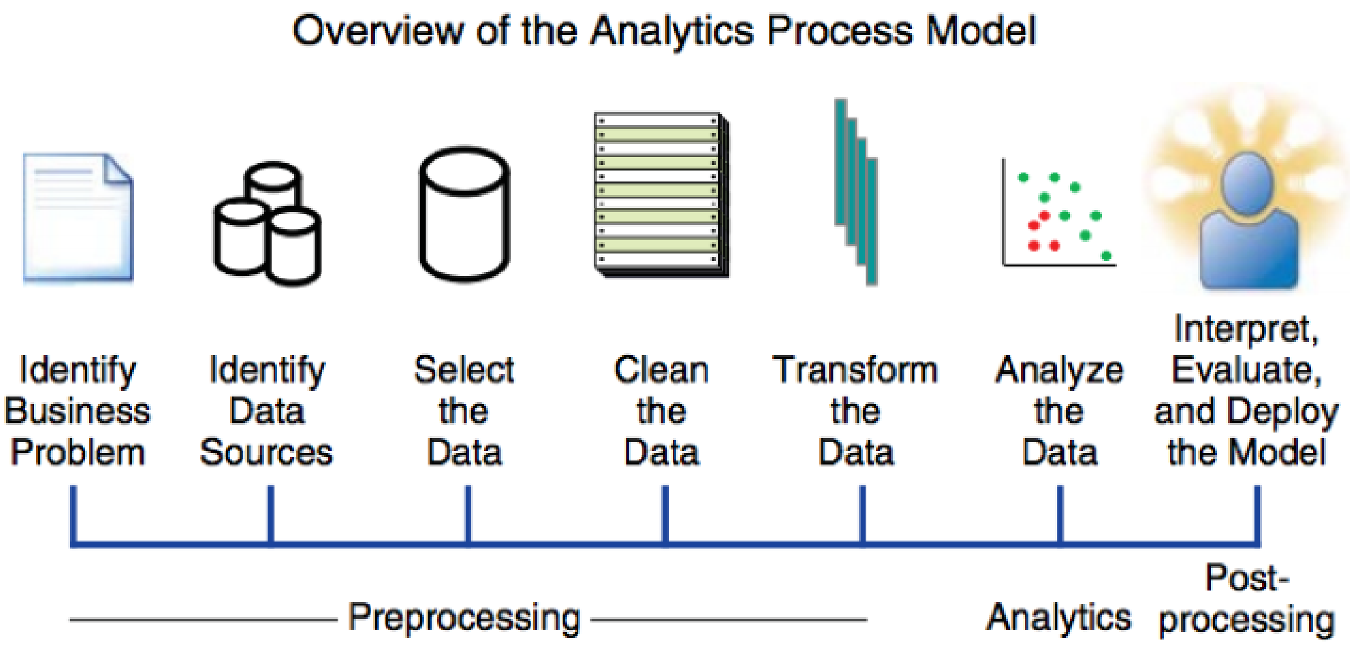


Big Data Analytics: Challenges

- Spatiotemporal correlation
- Scalability
- Missing data
- Bad data diversity
- Various types of uncertainties



Big Data Analytics: Challenges



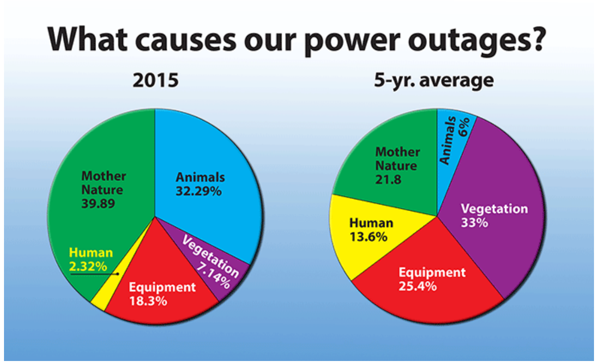
Applications: Outages caused by weather

- Background
- Transmission Insulator Failure Predictions
- Transmission Outage Prediction
- Prediction of Outages due to Distribution Vegetation
- Distribution Transformer Failure Prediction

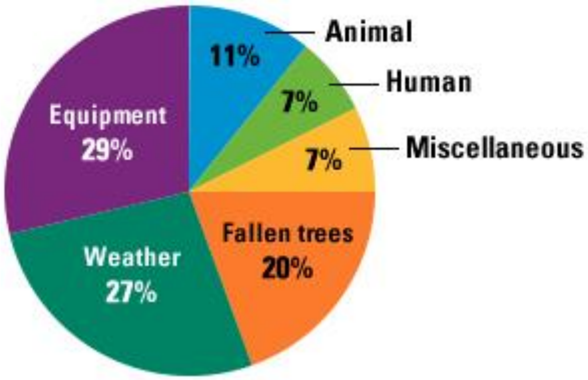
Background

P. Dehghanian, B. Zhang, T. Dokic and M. Kezunovic, “Predictive Risk Analytics for Weather-Resilient Operation of Electric Power Systems,” in IEEE Transactions on Sustainable Energy, Vol. 10, No., pp. 3-15, January 2019.

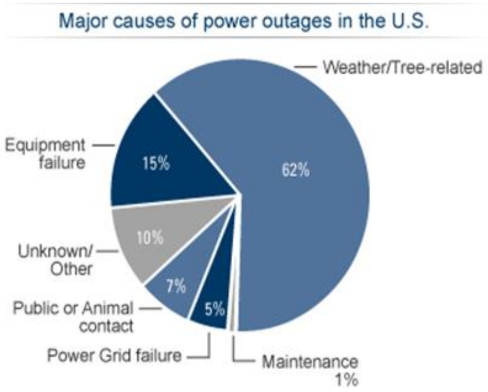
Causes

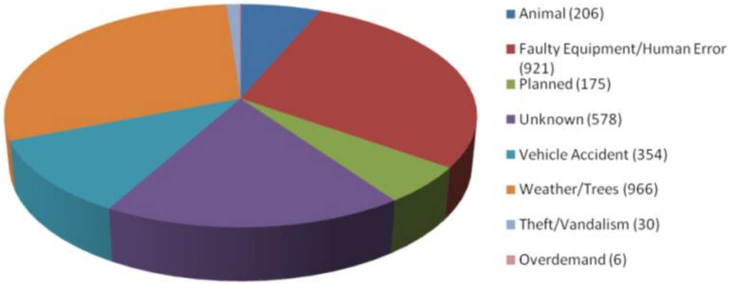


Source: Alaska Electric light and Power Company



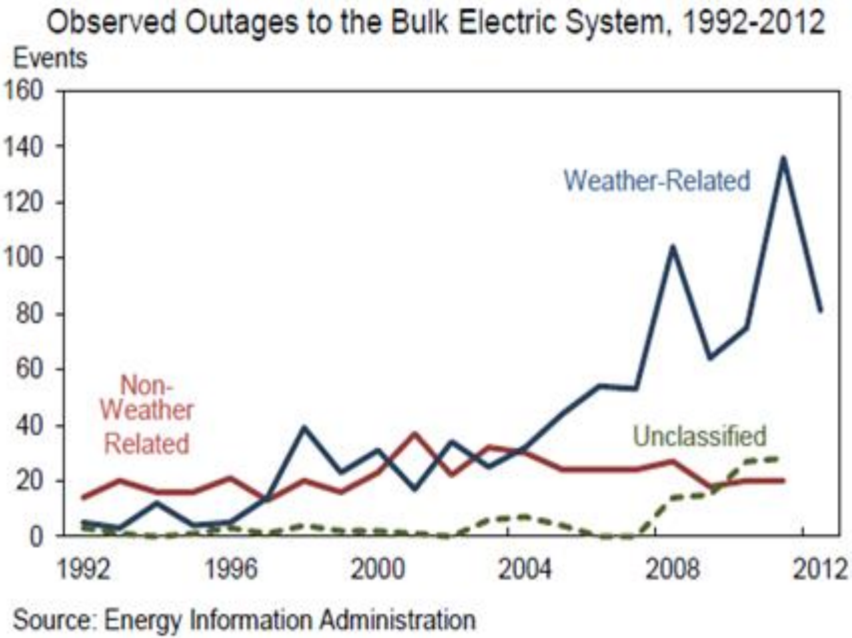
Source: We Energies

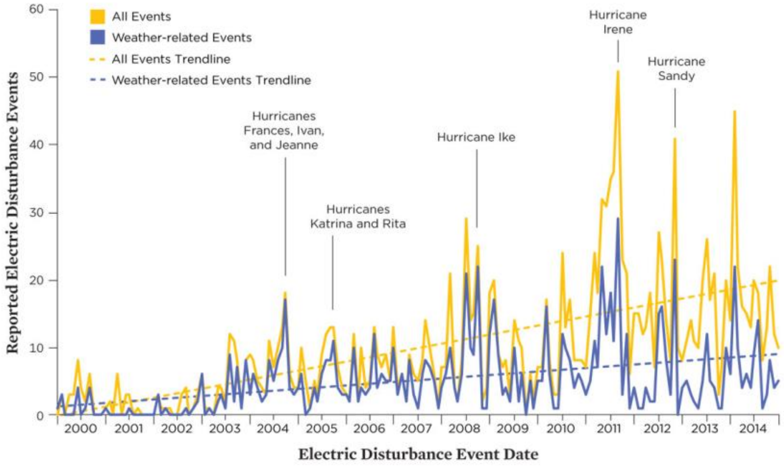




Source: Annual Eaton Investigation 2013

Manifestation



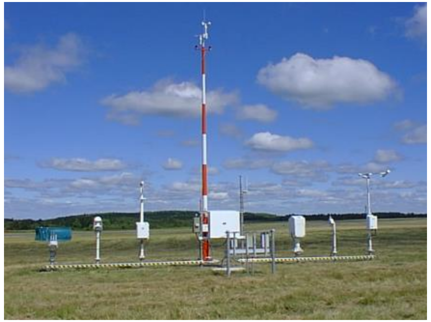


The Department of Energy tracks major electric disturbance events through Form OE-417. Utilities submit information about qualifying incidents, including when they occurred, where they occurred, what triggered them, and how many customers were affected. Notably, while the reported number of non-weather-related events is high, the vast majority of incidents resulting in customer outages occur because of weather.

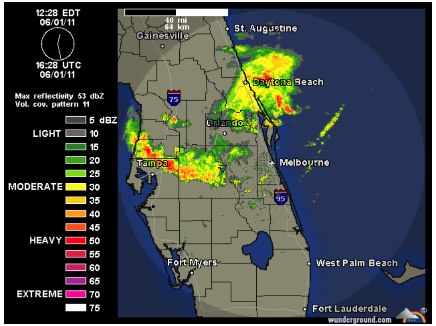
SOURCE: UCS ANALYSIS, BASED ON OE N.D.

Weather Data

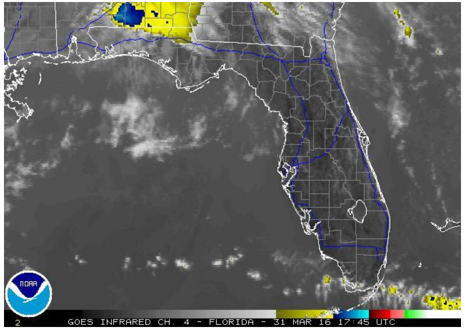
Weather Station



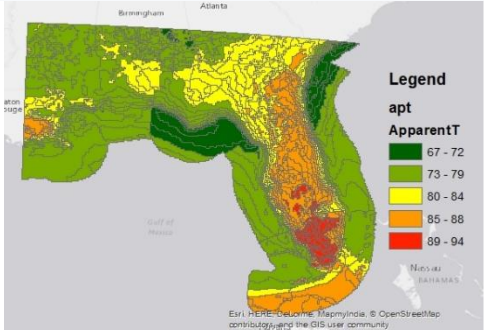
Radar



Satellite



National Digital Forecast Database (NDFD)



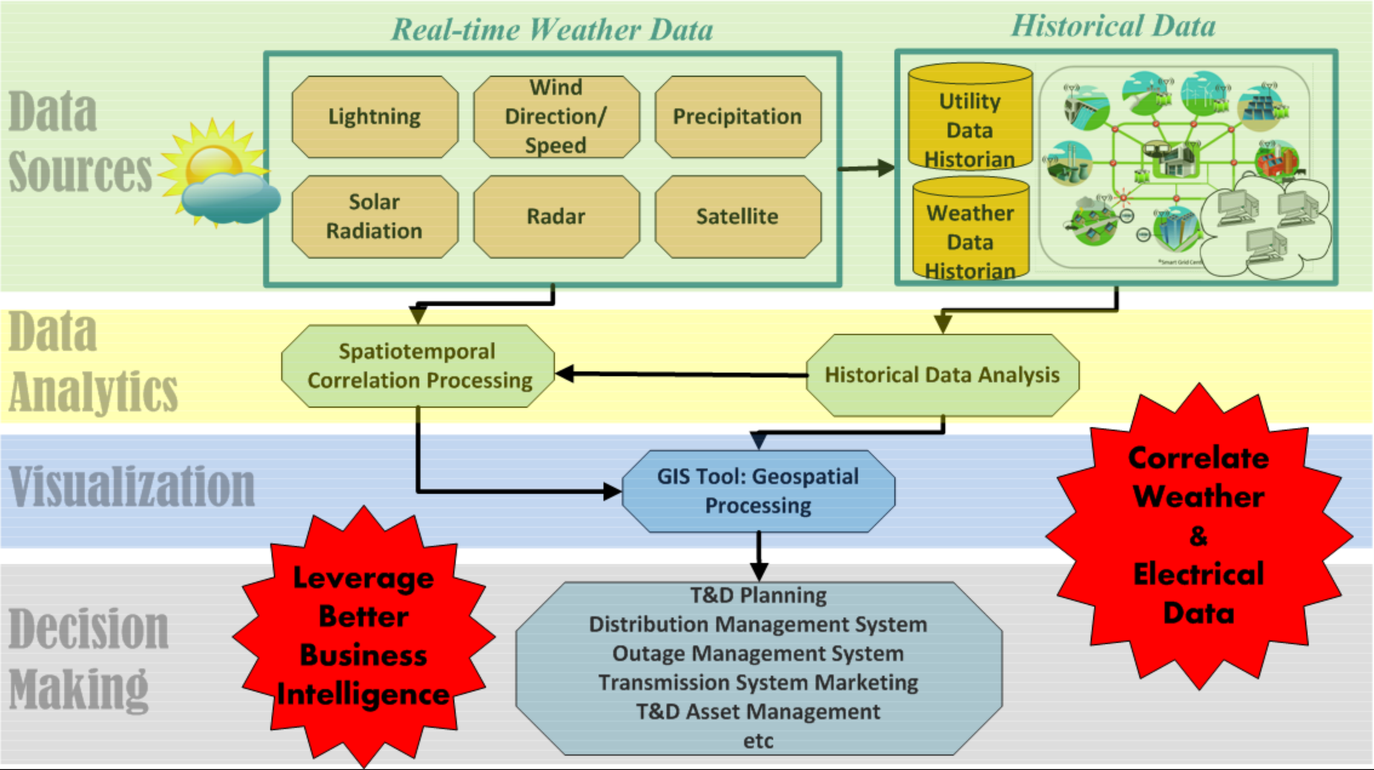
Example: Apparent Temperature
Data download: every 3 hours
Forecast for next 3 days
Data resolution: 3 hours

BD Data Properties

			VELOCITY	VOLUME	
	Source	Data Type	Temporal Resolution	Spatial Resolution	Measurements
VARIETY	Automated Surface Observing System	Land-Based	1 min	900 stations	Air Temperature, Dew Point, Relative Humidity, Wind Direction, Speed and Gust, Altimeter, Sea Level Pressure, Precipitation, Visibility…
	Level-2 Next Generation Weather Radar	Radar Data	5 min	160 high-resolution Doppler radar sites	Precipitation and Atmospheric Movement
	NOAA Satellite Database	Satellite Data	Hourly, daily, monthly	4 km	cloud coverage, hydrological observations (precipitation, cloud liquid water, total precipitable water, snow cover…), pollution monitoring…
	Vaisala U.S. National Lightning Detection Network	Lightning Data	Instantaneous	Median Location Accuracy 200-500m	Date and Time, Latitude and Longitude, Peak amplitude, Polarity, Type of event: Cloud or Cloud to Ground
	National Digital Forecast Database	Weather Forecast Data	3 hours	5 km	Wind Speed, Direction, and Gust, Relative Humidity, Convective Hazard Outlook, Tornado Probability, Probability of Thunderstorms…
	Texas Parks & Wildlife Department	Texas Ecological Mapping Systems Data	static	10 m	Distribution of different tree spices
	Texas Natural Resources Information System	NAIP	year	50 cm – 1 m	High Resolution Imagery
	National Aeronautics and Space Administration	3D Global Vegetation Map	static	1 km	Canopy height data
	National Cooperative Soil Survey	gSSURGO	static	10 m	Soil type
	Utility	Historical Outage Data	instantaneous	Feeder section	Location, start and end time and date, number of customers affected, cause code
		Tree Trimming Data	day	Feeder	Feeder location, date, trimming period, number of customers affected, cost of trimming
		Network GIS data	static	Infinity (shapefile)	Poles: location, material/class, height Feeders: location; conductor size, count, and material; nominal voltage
		Historical Maintenance Data	day	Tower location	Start and end date and time, location, type (maintenance, replacement), cost, number of customers affected
		Insulator asset data	static	Infinity (shapefile)	Surge Impedances of Towers and Ground Wires, Footing Resistance, Component BIL
		In-field measurements	instantaneous	Tower location	Leakage Current Magnitude, Flashover Voltage, Electric Field Distribution, Corona Discharge Detection, Infrared Reflection Thermography, Visual Inspection Reports

	Data Class	Data Source (Measurements)	VOLUME (Data file size)	VELOCITY (Rate of use)	VERACITY (Accuracy)
VARIETY	Utility measurements	SM	120GB per day/ device	Every 5-15 min	error <2.5%
		PMU	30GB per day/device	240 samples/sec	error <1%
		ICM	5GB per day/device	250 samples/sec	error <1%
		DFR	10MB per fault/device	1600 samples/sec	error <0.2%
	Weather data	Radar [27]	612 MB/day per radar	Every 4-10 min	1-2 dB; $m\ s^{-1}$
		Satellite [28]	At least 10 GB per day	Every 1-15 min	VIS<2%; IR<1-2K
		ASOS [29]	10 MB/day per station	Every 1 min	T-1.8°F, P<1%, Wind speed - 5%, RR - 4%
		NLDN [30]	40 MB/day	During lightning	SE < 200m, PCE <15%
		NDFD [31]	5-10 GB/day per model	1 - 12 hours	Varies by parameter
	Vegetation and Topography	TPWD EMST [32]	2.7 GB for Texas	static	SE < 10 m
		TNRIS [33]	300 GB for Texas	static	SE < 1 m
		LIDAR [34]	7 GB for Harris Co.	static	HE < 1m, VE < 150 cm

Implementation



Weather Driven Risk Analysis

Risk = Hazard x Vulnerability x Economic Impact

- Probability of hazardous weather conditions
- Depends on *Weather Forecast*
- Pick a moment in time (or a period of time) and estimate probability of hazardous conditions

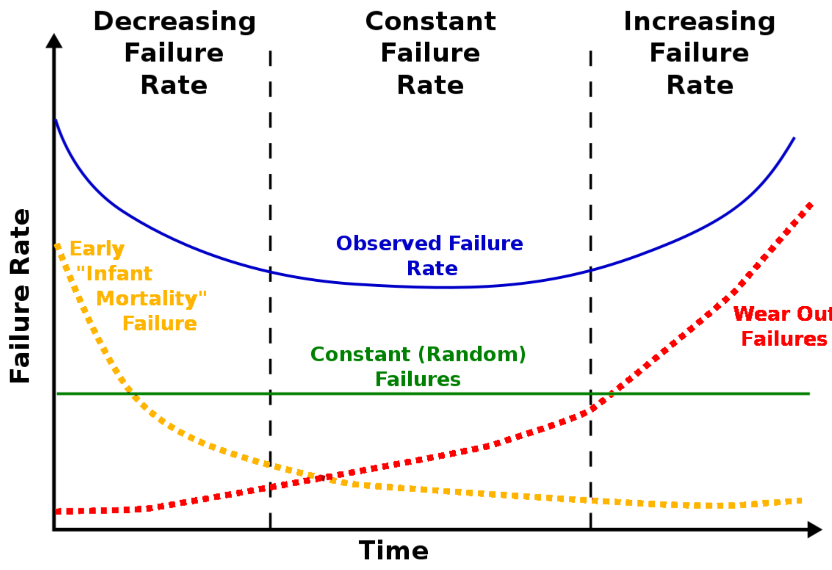
- Probability that hazardous conditions will cause an event in the network
- Depends on *Historical Weather and Outage Data*
- Learn from the historical data what may happen if hazardous conditions occur

- Expected economic impact in case of an event
- Depends on the type of economic loss that the user wants to consider
- Identify type of economic loss that is of interest for the study and calculate it

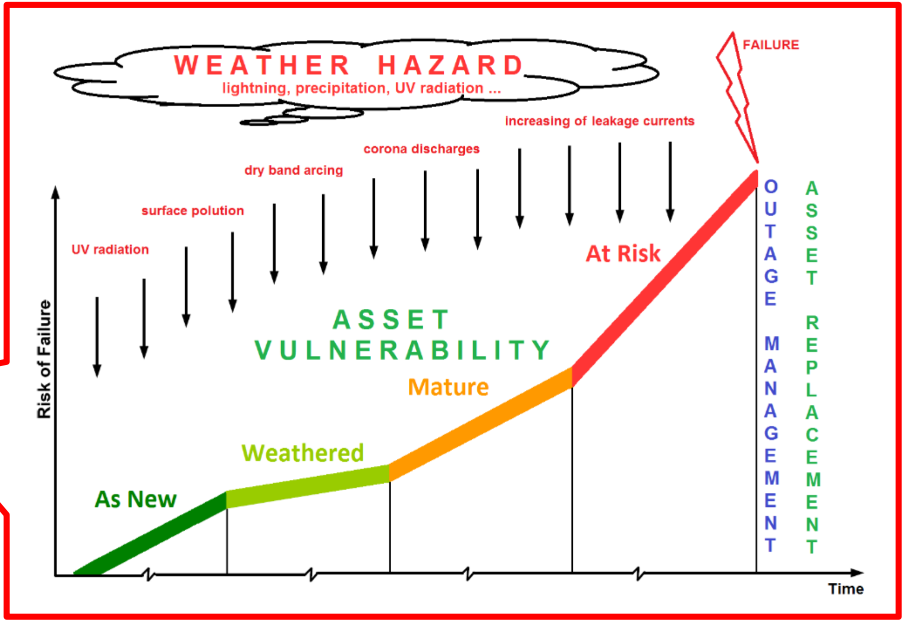
Transmission Insulation Coordination

M. Kezunovic, T. Dokic, R. Said, “Optimal Placement of Line Surge Arresters Based on Predictive Risk Framework Using Spatiotemporally Correlated Big Data,” at CIGRE General Session, Paris, France, Aug. 2018.

Asset Management for Insulators



*https://commons.wikimedia.org/wiki/File:Bathtub_curve.svg#/media/File:Bathtub_curve.svg



*A. Tzimas, et al. "Asset management frameworks for outdoor composite insulators." IEEE Transactions on Dielectrics and Electrical Insulation 19.6 (2012).

Risk Framework

(Risk = Hazard x Vulnerability)(Economic Impact)

$$R = P[T] \cdot P[C|T] \qquad u(C)$$

Intensity T – Lightning peak current

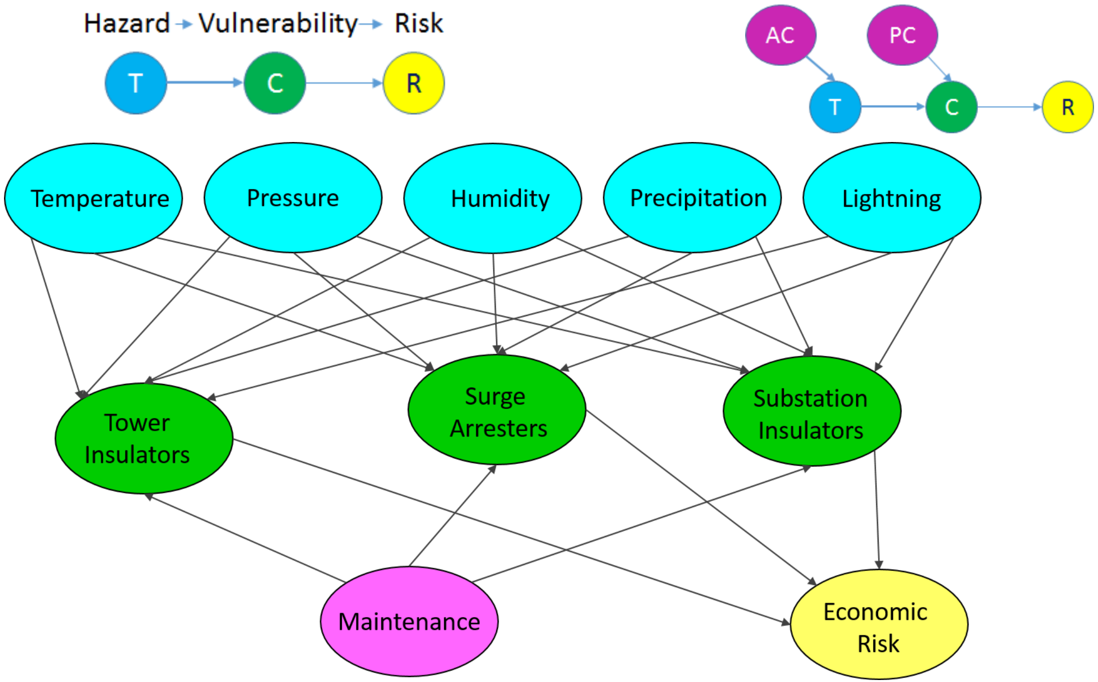
Hazard – Probability of a lightning strike

Vulnerability – Insulator vulnerability for a given hazard

Economic Impact – Estimated losses in case of insulator breakdown (cost of maintenance and replacement)

PSERC

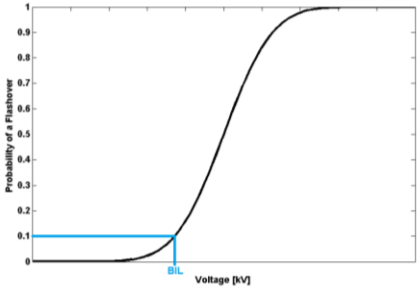
Insulator Risk Model



Modeling the Insulator BIL

Conventional method

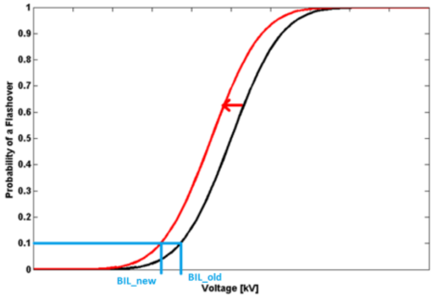
- BIL determined by insulator manufacturer.



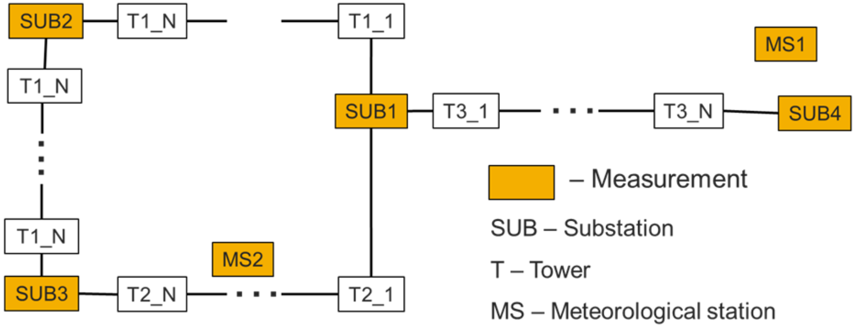
- Insulator breakdown probability determined statistically.
- Economic impact not taken into account.

BD approach

- Manufacturers standard BIL used only as a initial value. Standard BIL changes during the insulator lifetime.



- Insulator breakdown probability determined based on spatio-temporally referenced historical data and real-time weather forecast using data mining.
- Risk model includes economic impact in case of insulator breakdown.



Prediction Model

Lightning Detection Network

Weather Stations

Nodes: X = (Lightning Current, Temperature, Precipitation, Humidity, Pressure, BIL_old)

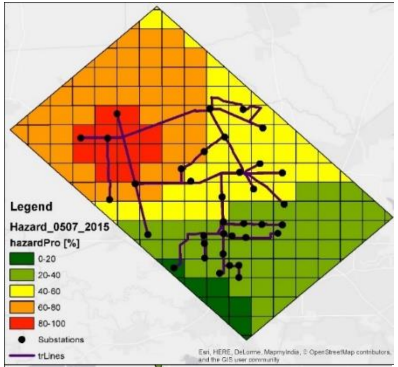
Y = (BIL_new)

Branches: Network Impedance Matrix

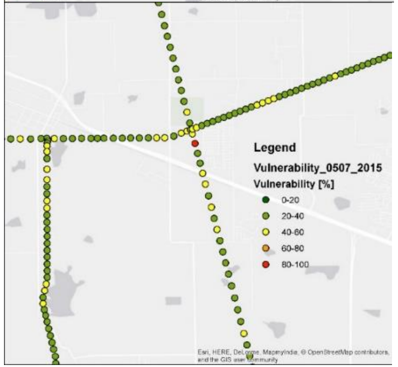
$$P(y|x) = \frac{1}{Z} exp\left(-\sum_{i=1}^{N} \sum_{k=1}^{K} \alpha_k \left(y_i - R_k(x)\right)^2 - \sum_{i,j} \sum_{l=1}^{L} \beta_l e_{ij}^{(l)} \, S_{ij}^{(l)}(x) \left(y_i - y_j\right)^2 \right)$$

Risk Maps

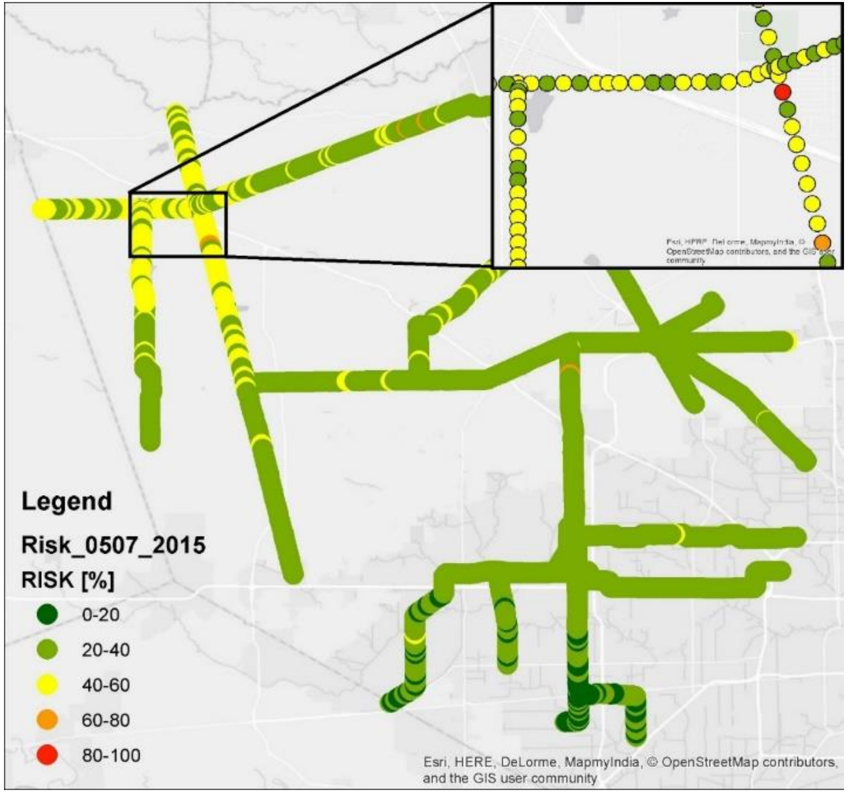
Weather Hazard



Network Vulnerability



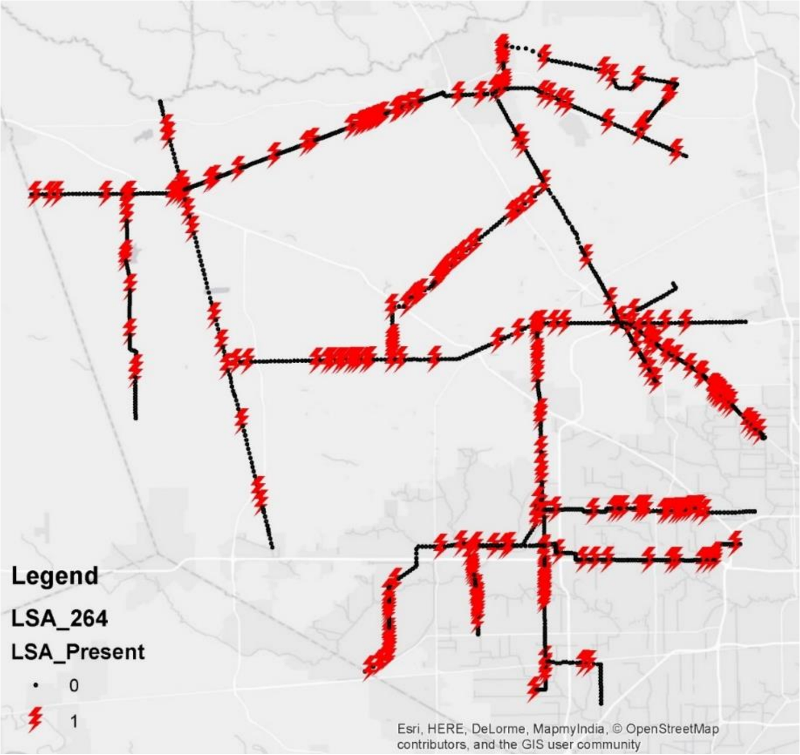
Risk Map



Optimal Location of LSAs

264 LSAs locations

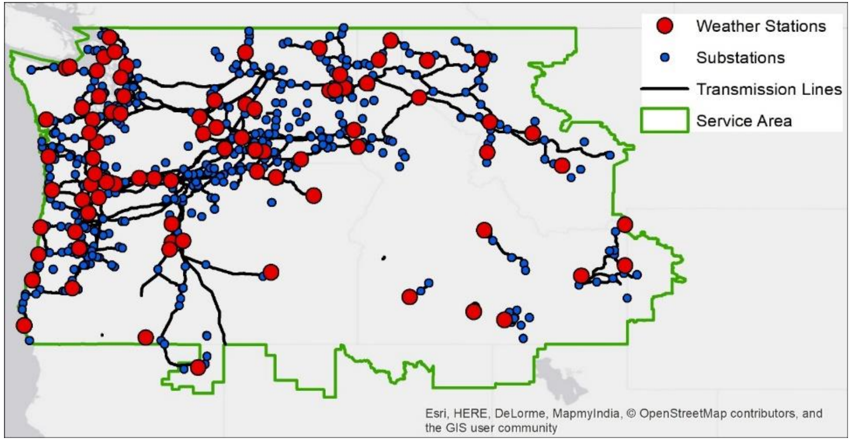
Total Risk Reduction = 72.69%



Transmission Outage Prediction

T. Dokic, M. Pavlovski, Dj. Gligorijevic, M. Kezunovic, Z. Obradovic, “Spatially Aware Ensemble-Based Learning to Predict Weather-Related Outages in Transmission,” The Hawaii International Conference on System Sciences – HICSS, Maui, Hawaii, January 2019.

Historical Weather Data

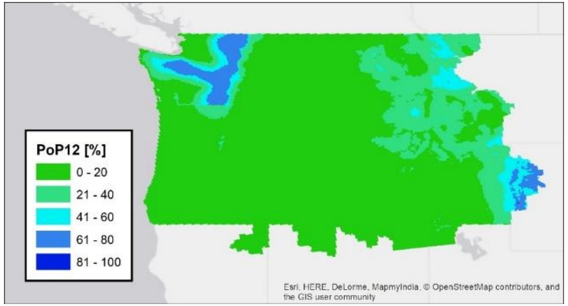


Fractions of missing data from ASOS observations

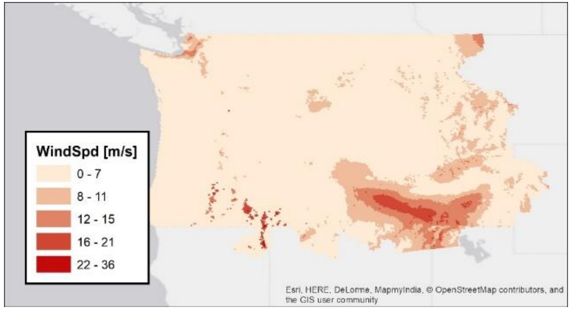
Temperature	Dew Point	Humidity	Wind Direction	Wind Speed	Precipitation	Pressure	Wind Gust	Weather Code
0.146	0.148	0.148	0.145	0.134	0.312	0.265	0.378	0.336

Weather Forecast

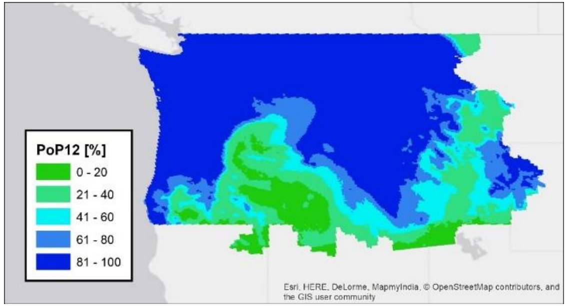
No Outage - Precipitation



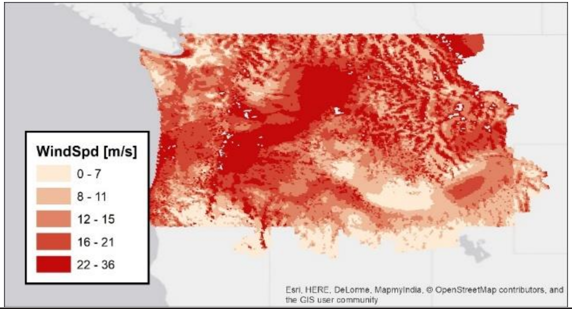
No Outage – Wind Speed



Outage - Precipitation



Outage – Wind Speed



Prediction Model

Estimate of an outage event with a probabilistic score (logistic regression)

- Various forecasted weather parameters as features related to an outage event
- A binary classification model to classify outage and no-outage events
- The output of an event is a probability score

Logistic regression

Logistic Regression [Kezunovic, *HICSS 2018*] (LR)

+ outputs outage probabilities
+ unconstrained convex optimization
– ignores spatial correlations between substations

Spatially-Aware Logistic Regression

+ embeds substations using [Newman, *PNAS 2006*]
+ accounts for spatial correlations
– cannot capture useful spatial substructures

Collaborative Logistic Ensemble Classifier (CLEC)

+ inspired by [Pavlovski, *IJCAI 2018*]

+ able to learn from spatial substructures

+ aims to account for generalization performance

Results – Outage Occurrence Prediction

Experimental Setup

- Training: data from 1999 to 2010
- Prediction horizon: 2010-2018
- Substations were embedded into a $\mathbf{50}$-dimenstional space based on their spatial proximity
- CLEC was run with $M = 30$ components
- $\eta = 30\%$ of the training data were sampled to construct the subset for each LR component

Model	Acc.	AUC	F1	Bias
LR	0.8467	0.9278	0.8097	0.6821
LR (spatial)	0.8624	0.9292	0.8242	0.7075
CLEC	**0.8919**	**0.9313**	**0.8532**	**0.7685**

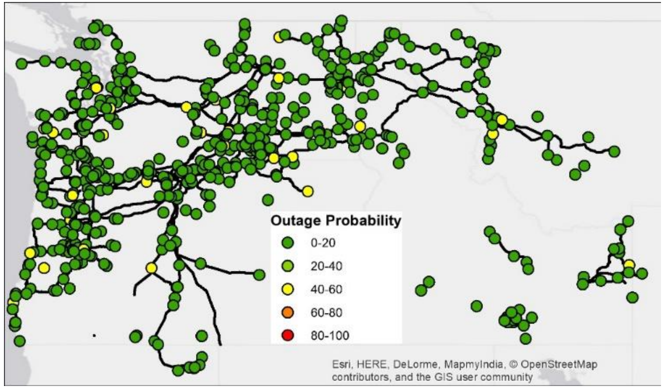
Prediction performance w.r.t. different evaluation metrics.

Discussion

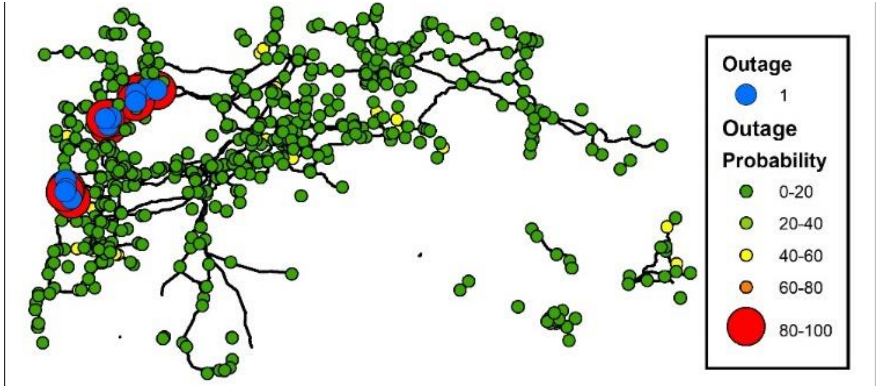
- LR (spatial) obtains greater classification performance compared to LR
  - ⇨ supports the hypothesis **that spatial information is truly relevant for this task**
- CLEC outperforms its alternatives, yielding higher values for accuracy, AUC and F1
- Large lift in Bias
  - ⇨ shows the **benefit of using a subsampling-based ensemble scheme**

Risk Maps

No Outage



Multiple Lightning Outage



- **No outages occurred** ⇨ outage probabilities are **smaller than 60%** for all substations
- Outages occurred ⇨ the area around the outages has points with probability over 80%
- **In general**: Both LR and CLEC do well at guessing the areas of outage occurrences
- **However**: CLEC does better than LR making prediction **more precise on a spatial level**
  - ⇨ the number of high risk areas far away from the outage locations is much smaller

Distribution Vegetation Management

T. Dokic, M. Kezunovic, “Predictive Risk Management for Dynamic Tree Trimming Scheduling for Distribution Networks,” IEEE Transactions on Smart Grid, Vol. 10, No. 5, pp. 4776-4785, September 2018.

Risk Framework

(Risk = Hazard x Vulnerability)(Economic Impact)

$$R = P[T] \cdot P[C|T] \qquad u(C)$$

Intensity T – Weather severity

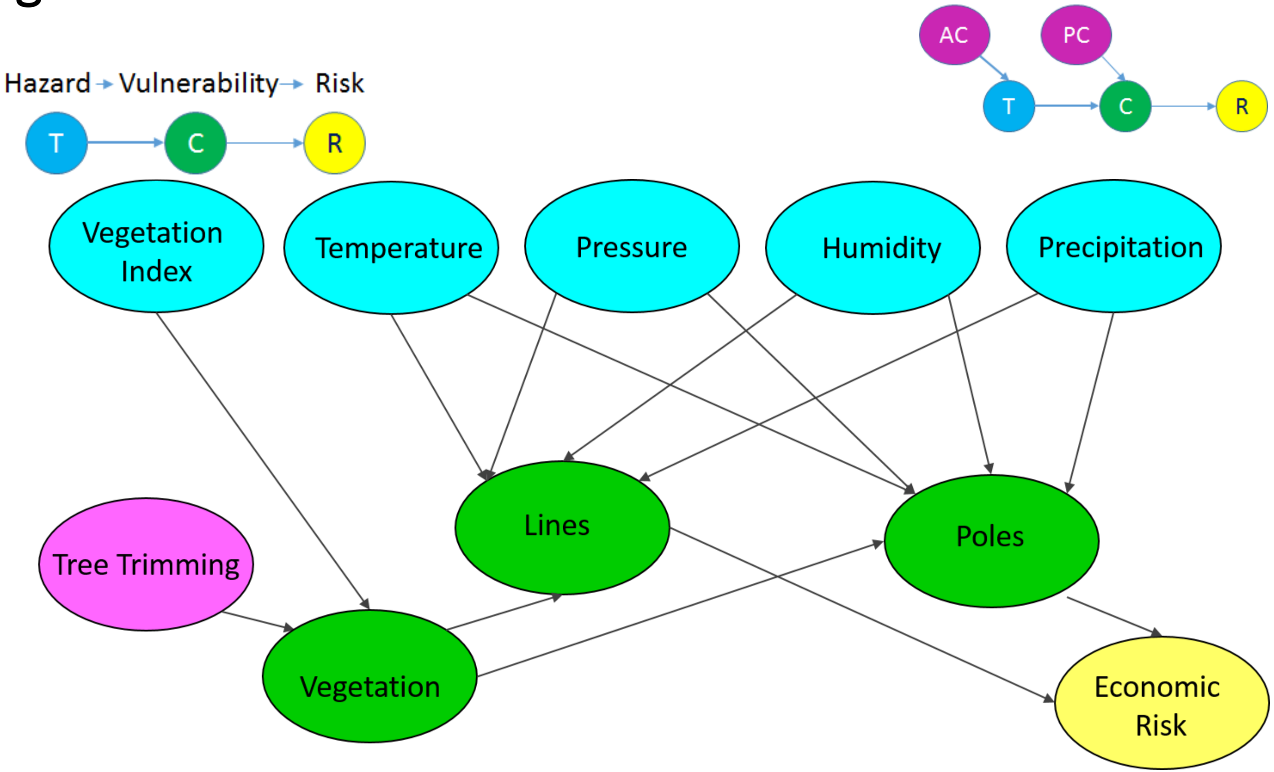
Hazard – Probability of a severe weather impact

Vulnerability – Probability of a vegetation caused outage for a given hazard

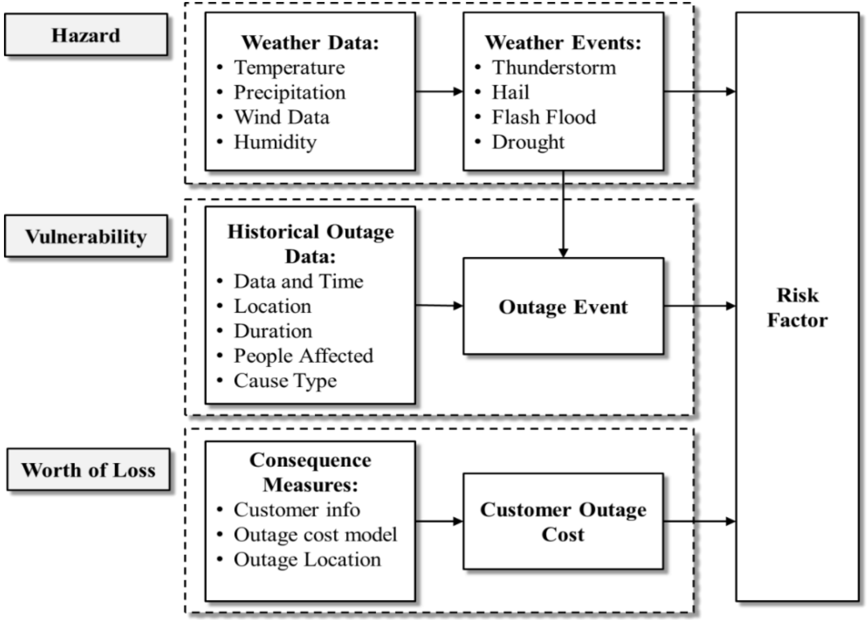
Economic Impact – Cost of periodic and reactive tree trimming

PSERC

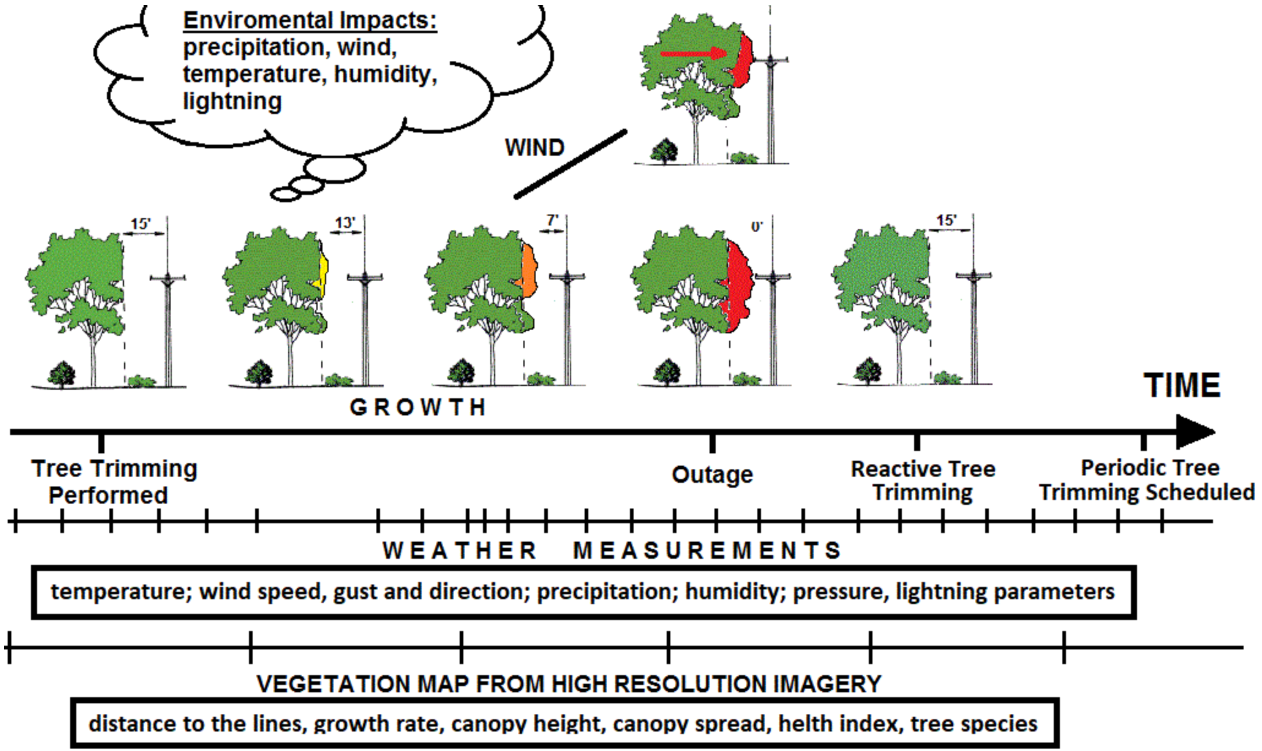
Vegetation Risk Model



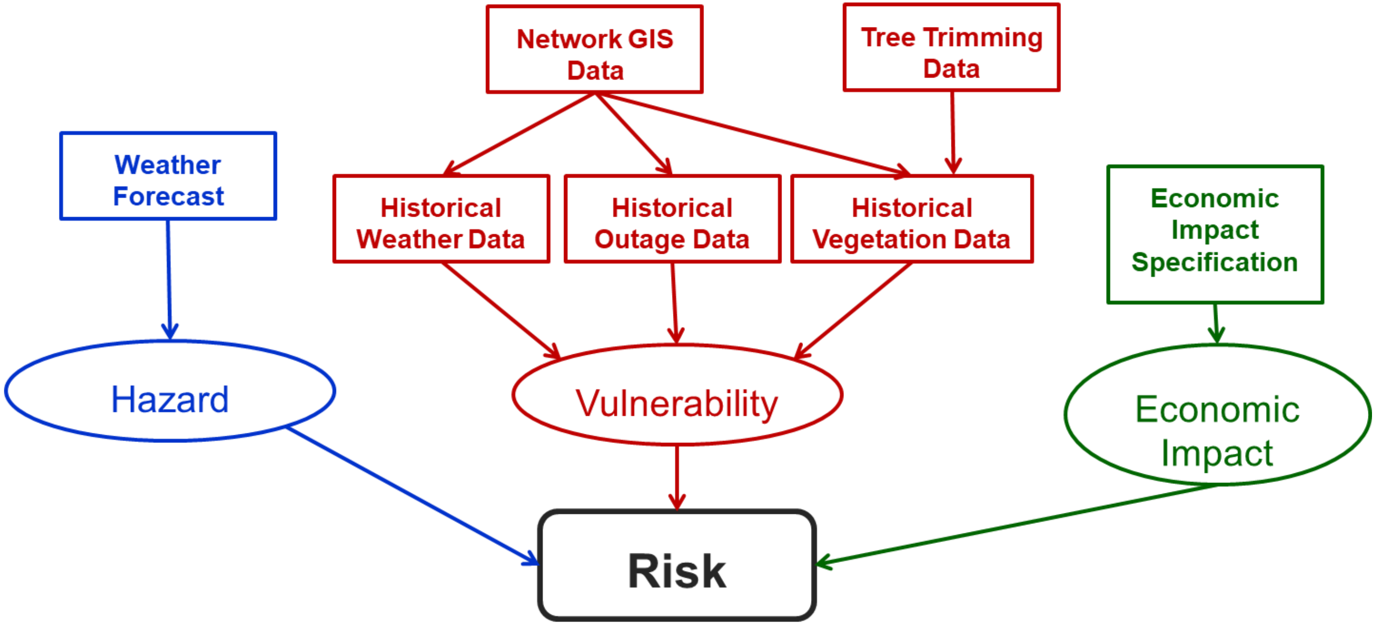
BD use in modeling weather Impacts



Big Data for Vegetation Management



Risk Model



Network GIS Data
Tree Trimming Data
Weather Forecast
Historical Weather Data
Historical Outage Data
Historical Vegetation Data
Economic Impact Specification
Hazard
Vulnerability
Economic Impact
Risk

Optimal Tree Trimming

For total of N feeder sections maximize the reduction in vegetation risk

$$max\left\{R=\sum_{t=1}^{T}\frac{1}{N}\sum_{n=1}^{N}\Delta R_{n,t}\cdot F_{n,t}\right\}$$

$$F_{n,t}=\begin{cases}0, & \text{section not trimmed}\\ 1, & \text{section trimmed}\end{cases}$$

Difference in component risk before and after action:

$$\Delta R_{n,t}=R_{n,t}^{before}-R_{n,t}^{after}$$

Constrains:

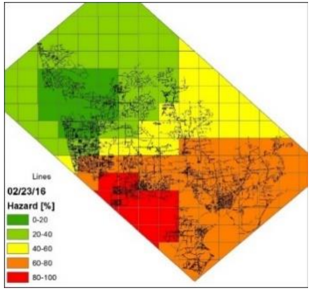
Total cost of tree trimming limit: $\sum_{t=1}^{T}\sum_{n=1}^{N}F_{n,t}\cdot PC_{n,t}\leq PA$

One section trimmed at the time: For $t=1,\ldots,T$, $\sum_{n=1}^{N}F_{n,t}=1$

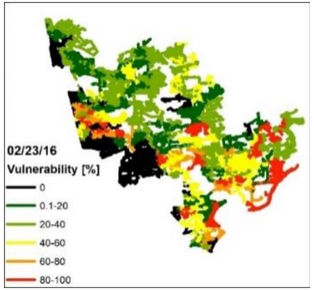
n	feeder section
N	total number of feeder sections
t	time instance
T	number of time instances
$\Delta R_{n.t}$	reduction in risk after trimming
$PC_{n,t}$	cost of trimming on one section
PA	total allocated tree trimming budget

Risk Maps

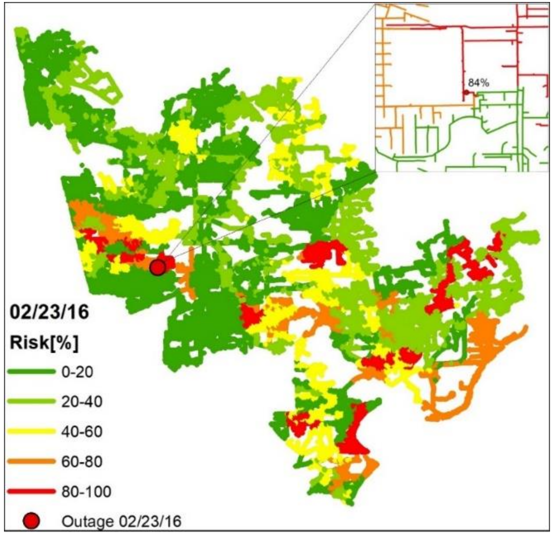
Weather Hazard



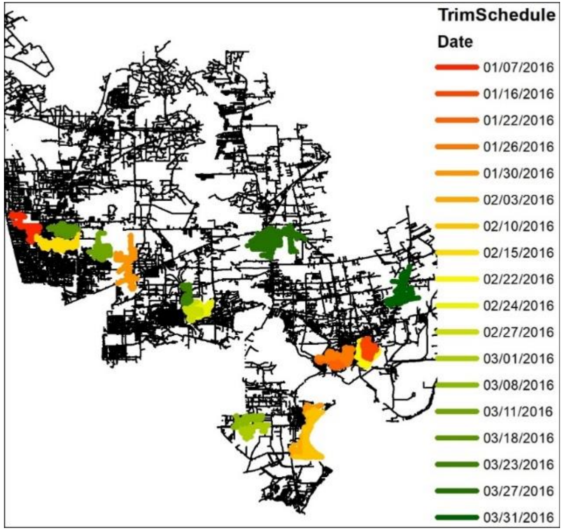
Network Vulnerability



Risk Map



Optimal Tree Trimming Schedule



Overall risk reduction 32.85%

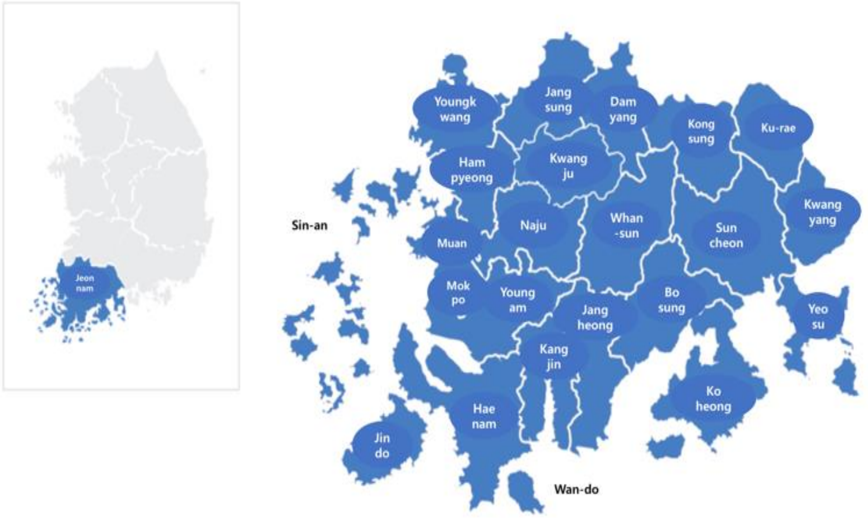
Reactive tree trimming cost reduction 27.2%

Distribution Transformer Failure Prediction

E. Hui Ko, T. Dokic, M. Kezunovic, “Prediction Model for the Distribution Transformer Failure using Correlation of Weather Data,” CIGRE 5th International Colloquium Transformer Research and Asset Management, Opatija, Croatia, October 2019.

The System – South Korea

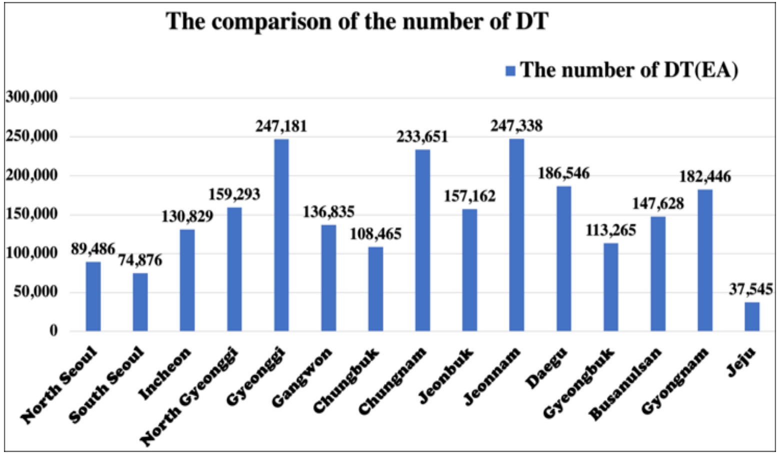
JeonllaNam-do Area



Distribution Facilities in JeonllaNam-do Area

Transformer			Protective Devices		
Bank	Number	Capacity (kVA)	Breaker	Equipment	COS
104	243	9,252	12	1.4	72

The comparison of the number of DT in South Korea

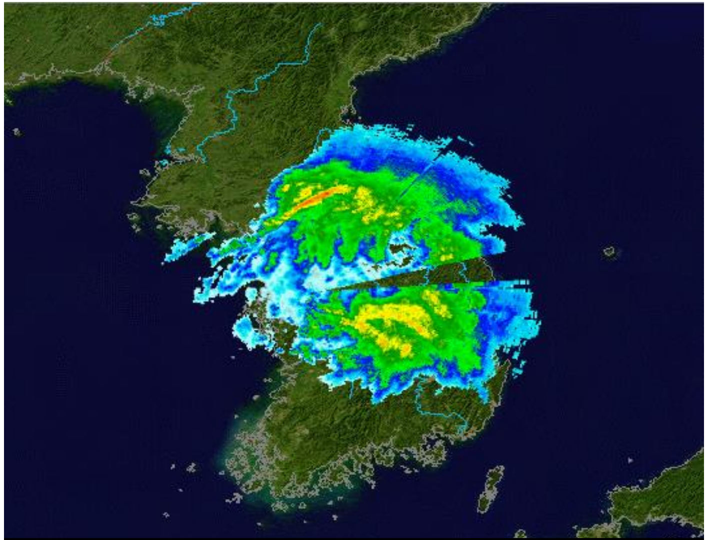


Transformer Failures

Cause	The number	Rate [%]
Aging	337	27.2
Weather	333	26.9
Corrosion	155	12.5
Animal contact	145	11.7
Out of maintenance	92	7.4
Overloading	87	7.0
Object contact	25	2.0
Tree	18	1.5
People error	16	1.3
Manufacture error	15	1.2
Flooding	13	1.0
Installment error	3	0.2
Fire	1	0.1
Total	1,240	100

Causes of DT Failures in JeonllaNam-do of South Korea (2011 - 2018)

Weather Data

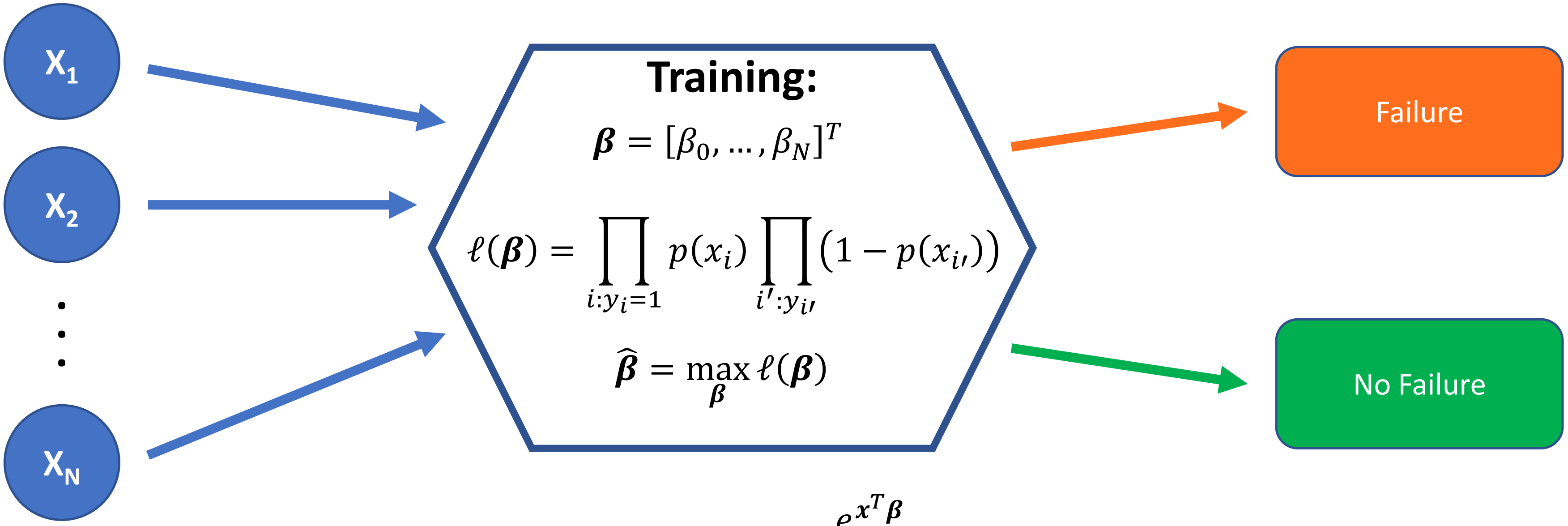


NOAA/U.S. Air Force, Blue Marble map
https://www.ncdc.noaa.gov/nexradinv/chooseday.jsp?id=rksg

Historical Weather Measurements

Lightning [0/1] (LI)	Average Temperature [˚F] (AT)	Highest Temperature [˚F] (HT)	Relative Humidity [%] (RH)	Maximum Wind Speed [m/s] (MWS)	Wind Gust [m/s] (WG)

Logistic Regression



Testing: $p(\boldsymbol{x}) = \dfrac{e^{\boldsymbol{x}^T \boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}^T \boldsymbol{\beta}}}$

Evaluation Setup

- Failure data spans from 2011 to 2018.
- The total number of DT failure is 237.
- Data divided into the training (90% of data) and testing (10% of data) sets.
- Total of 148 of no failure cases added.
- Total number of events is 237 + 148 = 385.
- The degree of high temperature (HT) is classified into three temperature thresholds: 82.4°F, 86°F, and 89.6°F in order to make interpretation of HT coefficient precise.

The true positive rate of a classifier:

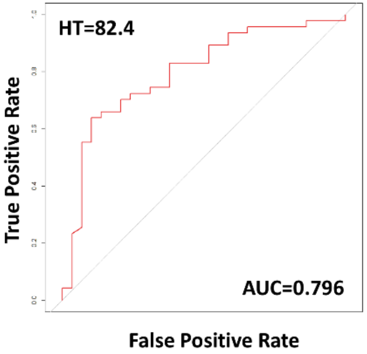
$$\text{tp rate} \approx \frac{Positive\ correctly\ classified}{Total\ positives}$$

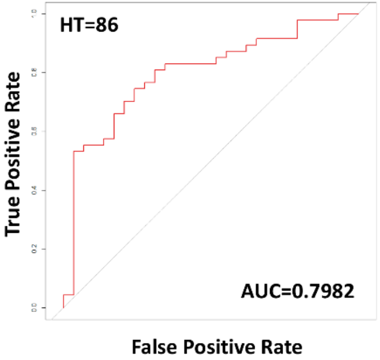
The false positive rate of the classifier:

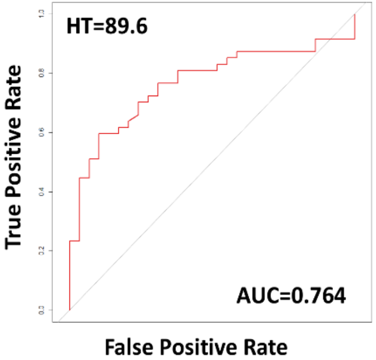
$$\text{fp rate} \approx \frac{Negatives\ incorrectly\ classified}{Total\ negatives}$$

Prediction Results

Receiver Operating Characteristics Curve (ROC)







Event vs. Prediction of Failure

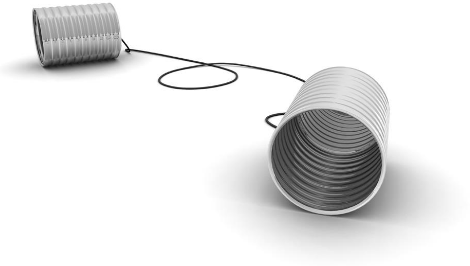
Event	Failure (Y/N)	Prediction	
		Y=0	Y=1
HT 86°F or below	Y=0	113	47
	Y=1	35	190
HT 86°F - 89.6°F	Y=0	112	47
	Y=1	36	190
HT 89.6°F or above	Y=0	111	54
	Y=1	37	183

Conclusions

- The **weather impacts on outage** is prominent through wear and tear, as well as from stresses due to impacts of storms
- The **spatial granularity of prediction** and **localization of outages** can be improved by embeddings and **modeling of spatial interactions.**
- Type of prediction algorithm depends on the application: **Linear Regression, Logistic Regression**
- A **real-time mapping system is needed** to observe **Risk**
- **Predictive system** allows for more **<u>proactive and cost-effective</u>** outage management, asset management, and operation.

Thank you!

Questions?



Mladen Kezunovic
Tel: (979) 845-7509
E-mail: kezunov@ece.tamu.edu

Together -
building a prosperous future

where energy is
clean, abundant, reliable, safe, secure and affordable